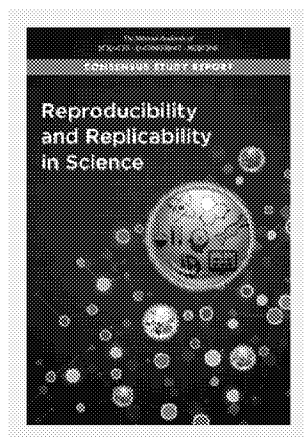


This PDF is available at <http://nap.edu/25303>

SHARE



Reproducibility and Replicability in Science (2019)

DETAILS

256 pages | 6 x 9 | PAPERBACK

ISBN 978-0-309-48616-3 | DOI 10.17226/25303

CONTRIBUTORS

Committee on Reproducibility and Replicability in Science; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on National Statistics; Division of Behavioral and Social Sciences and Education; Nuclear and Radiation Studies Board; Division on Earth and Life Studies; Board on Mathematical Sciences and Analytics; Committee on Applied and Theoretical Statistics; Division on

SUGGESTED CITATION Sciences; Board on Research Data and Information;

Committee on Science, Engineering, Medicine, and Public Policy; Policy and National Academies of Sciences, Engineering, and Medicine 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press.

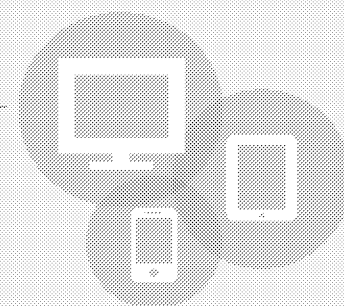
<https://doi.org/10.17226/25303>.

GET THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

Copyright © National Academy of Sciences. All rights reserved.

Summary

One of the pathways by which scientists confirm the validity of a new finding or discovery is by repeating the research that produced it. When a scientific effort fails to independently confirm the computations or results of a previous study, some argue that the observed inconsistency may be an important precursor to new discovery while others fear it may be a symptom of a lack of rigor in science. When a newly reported scientific study has far-reaching implications for science or a major potential impact on the public, the question of its reliability takes on heightened importance. Concerns over reproducibility and replicability have been expressed in both scientific and popular media.

As these concerns increased in recent years, Congress directed the National Science Foundation (NSF) to contract with the National Academies of Sciences, Engineering, and Medicine to undertake a study to assess reproducibility and replicability in scientific and engineering research and to provide findings and recommendations for improving rigor and transparency in research.

THE ROLE OF REPRODUCIBILITY AND REPLICABILITY IN SCIENCE

To gain knowledge about the world and to seek new discoveries through scientific inquiry, scientists often first perform exploratory research. This kind of work is only the start toward establishing new knowledge. The path from a new discovery reported by a single scientist (or single group of scientists) to adoption by others involves confirmatory research (i.e., testing

and confirmation), an examination of the limits of the original result (by the original researchers or others), and development of new or expansion of existing scientific theory. This process may confirm and extend existing knowledge, or it may upend previous knowledge and replace it with more accurate scientific understanding of the natural world. The scientific enterprise depends on the ability of the scientific community to scrutinize scientific claims and to gain confidence over time in results and inferences that have stood up to repeated testing.

Important throughout this process is the sharing of data and methods and the estimation, characterization, and reporting of uncertainty. Reporting of uncertainty in scientific results is a central tenet of the scientific process, and it is incumbent on scientists to convey the appropriate degree of uncertainty to accompany original claims.

Because of the intrinsic variability of nature and limitations of measurement devices, results are assessed probabilistically, with the scientific discovery process unable to deliver absolute truth or certainty. Instead, scientific claims earn a higher or lower likelihood of being true depending on the results of confirmatory research. New research can lead to revised estimates of this likelihood.

DEFINITIONS

The terms reproducibility and replicability have different meanings and uses across science and engineering, which has led to confusion in collectively understanding problems in reproducibility and replicability. The committee adopted specific definitions for the purpose of this report to clearly differentiate between the terms, which are otherwise interchangeable in everyday discourse.

Reproducibility is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis. This definition is synonymous with “computational reproducibility,” and the terms are used interchangeably in this report.

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. Two studies may be considered to have replicated if they obtain consistent results given the level of uncertainty inherent in the system under study.

Generalizability, another term frequently used in science, refers to the extent that results of a study apply in other contexts or populations that

differ from the original one.¹ A single scientific study may include elements or any combination of these concepts.

In short, reproducibility involves the *original* data and code; replicability involves *new* data collection to test for consistency with previous results of a similar study. These two processes also differ in the type of results that should be expected. In general, when a researcher transparently reports a study and makes available the underlying digital artifacts, such as data and code, the results should be computationally reproducible. In contrast, even when a study was rigorously conducted according to best practices, correctly analyzed, and transparently reported, it may fail to be replicated.

REPRODUCIBILITY

The committee's definition of reproducibility is focused on computation because of its major and increasing role in science. Most scientific and engineering research disciplines use computation as a tool. The abundance of data and widespread use of computation have transformed many disciplines, but this revolution is not yet uniformly reflected in how scientists develop and use software and how scientific results are published and shared. These shortfalls have implications for reproducibility, because scientists who wish to reproduce research may lack the information or training they need to do so.

When results are produced by complex computational processes using large volumes of data, the methods section of a scientific paper is insufficient to convey the necessary information for others to reproduce the results. Additional information related to data, code, models, and computational analysis is needed for others to computationally reproduce the results.

RECOMMENDATION 4-1: To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis, unless such information is restricted by nonpublic data policies. That information should include the data, study methods, and computational environment:

- the input data used in the study either in extension (e.g., a text file or a binary) or in intension (e.g., a script to generate the data), as well as intermediate results and output data for steps that are nondeterministic and cannot be reproduced in principle;

¹ The same definition of generalizability as used by NSF (Bollen et al., 2015).

- a detailed description of the study methods (ideally in executable form) together with its computational steps and associated parameters; and
- information about the computational environment where the study was originally executed, such as operating system, hardware architecture, and library dependencies. (Library dependency,² in the context of research software as used here, is the relationship of pieces of software that are needed for another software to run. Problems often occur when installed software has dependencies on specific versions of other software.)

Some fields of scientific inquiry, such as geoscience, involve complex data gathering from multiple sensors, modeling, and algorithms that cannot all be readily captured and made available for other investigators to reproduce. Some research involves nonpublic information that cannot legally be shared, such as patient records or human subject data. Other research may involve instrumentation with internal data processing algorithms that are not directly accessible to the investigator due to proprietary restrictions. The committee acknowledges such circumstances. However, when feasible to collect and share the necessary information, computational results are expected to be reproducible.

Expected Results from Attempts to Reproduce Research

If sufficient data, code, and methods description are available and a second researcher follows the methods described by the first researcher, one expects in many cases full bitwise reproduction of the original results—that is, obtaining the same exact numeric values. For some research questions, bitwise reproducibility may be relaxed and reproducible results could be obtained within an accepted range of variation. Understanding the range of variation and the limits of computational reproducibility in increasingly complex computational systems, such as artificial intelligence, high-performance computing, and deep learning, is an active area of research.

RECOMMENDATION 4-2: The National Science Foundation should consider investing in research that explores the limits of computational reproducibility in instances in which bitwise reproducibility is not reasonable in order to ensure that the meaning of consistent computational results remains in step with the development of new computational hardware, tools, and methods.

² This definition was corrected during copy editing between release of the prepublication version and this final, published version.

Exact reproducibility does not guarantee the correctness of the computation. For example, if an error in code goes undetected and is reapplied, the same erroneous result may be obtained.

The Extent of Non-Reproducibility in Research

Reproducibility studies can be grouped into one of two kinds: (1) direct, which regenerate computationally consistent results; and (2) indirect, which assess the transparency of available information to allow reproducibility.

Direct assessments of reproducibility, replaying the computations to obtain consistent results, are rare in comparison to indirect assessments of transparency, that is, checking the availability of data and code. Direct assessments of computational reproducibility are more limited in breadth and often take much more time and resources than indirect assessments of transparency.

The standards for success of direct and indirect computational reproducibility assessments are neither universal nor clear-cut. Additionally, the evidence base of computational non-reproducibility³ across science is incomplete. Thus, determining the extent of issues related to computational reproducibility across fields or within fields of science would be a massive undertaking with a low probability of success. Notably, however, a number of systematic efforts to reproduce computational results across a variety of fields have failed in more than one-half of the attempts made, mainly due to insufficient detail on digital artifacts, such as data, code, and computational workflow.

REPLICABILITY

Unlike the typical expectation of reproducibility between two computations, expectations about replicability are more nuanced. A successful replication does not guarantee that the original scientific results of a study were correct, nor does a single failed replication conclusively refute the original claims. Furthermore, a failure to replicate can be due to any number of factors, including the discovery of new phenomena, unrecognized inherent variability in the system, inability to control complex variables, and substandard research practices, as well as misconduct.

³ “Non-reproducible” and “irreproducible” are both used in scientific work and are synonymous.

The Extent of Non-Replicability in Research

The committee was asked to assess what is known about the extent of non-replicability in science and, if necessary, to identify areas that may need more information to ascertain it. One challenge in assessing the extent of non-replicability across science is that different types of scientific studies lead to different or multiple criteria for determining a successful replication. The choice of criteria can affect the apparent rate of non-replication and calls for judgment and explanation. Therefore, comparing results across replication studies may be compromised because different replication studies may test different study attributes and rely on different standards and measures for a successful replication.

Another challenge is that there is no standard across science for assessing replication between two results. The committee outlined a number of criteria central to such comparisons and highlights issues with misinterpretation of replication results using statistical inference. A number of parametric and nonparametric methods may be suitable for assessing replication across studies. However, it is restrictive and unreliable to accept replication only when the results in both studies have attained “statistical significance,” that is, when the p -values in both studies have exceeded a selected threshold. Rather, in determining replication, it is important to consider the distributions of observations and to examine how similar these distributions are. This examination would include summary measures, such as proportions, means, standard deviations (uncertainties), and additional metrics tailored to the subject matter.

The issue of uncertainty merits particular attention. Scientific studies have irreducible uncertainties, whether due to random processes in the system under study, limits to scientific understanding or ability to control that system, or limitations in the precision of measurement. It is the job of scientists to identify and characterize the sources of uncertainty in their results. Quantification of uncertainty allows scientists to compare their results (i.e., to assess replicability), identify contributing factors and other variables that may affect the results, and assess the level of confidence one should have in the results. Inadequate consideration of these uncertainties and limitations when designing, conducting, analyzing, and reporting the study can introduce non-replicability.

RECOMMENDATION 5-1: Researchers should, as applicable to the specific study, provide an accurate and appropriate characterization of relevant uncertainties when they report or publish their research. Researchers should thoughtfully communicate all recognized uncertainties and estimate or acknowledge other potential sources of uncertainty that bear on their results, including stochastic uncertainties and

uncertainties in measurement, computation, knowledge, modeling, and methods of analysis.

An added challenge in assessing the extent of non-replicability is that many replication studies are not reported. Because many scientists routinely conduct replication tests as part of a follow-on experiment and do not report replication results separately, the evidence base of non-replicability across all science and engineering research is incomplete.

Finally, non-replicability may be due to multiple sources, some of which are beneficial to the progression of science, and some of which are not. The overall extent of non-replicability is an inadequate indicator of the health of science.

Recognizing these limitations, the committee examined replication studies in the natural and clinical sciences (e.g., general biology, genetics, oncology, chemistry) and social sciences (e.g., economics, psychology) that report frequencies of replication ranging from fewer than one of five studies to more than three of four studies.

Sources of Non-Replicability in Research

In an attempt to tease apart factors that contribute to non-replicability, the committee classified sources of non-replicability into those that are potentially helpful to gaining knowledge and those that are unhelpful.

Potentially helpful sources of non-replicability. Potentially helpful sources of non-replicability include inherent but uncharacterized uncertainties in the system under study. These sources are a normal part of the scientific process, due to the intrinsic variation and complexity of nature, scope of current scientific knowledge, and limits of our current technologies. They are not indicative of mistakes; rather, they are consequences of studying complex systems with imperfect knowledge and tools.

These sources also include deliberate choices made by researchers that may increase the occurrence of non-replicable results. For example, reasonable decisions made by one researcher on the cleaning of a data collection may result in a different final dataset that would affect the study's results. Or a study that has a higher chance of discovering new effects may also have a higher chance of producing non-replicable results due to unknown aspects of the system and methods used in the discovery. Researchers may choose to accept a higher false-positive rate for initial (i.e., exploratory) research. A researcher may also opt to allow some potential sources of non-replicability—for example, a lower number of study participants—because of considerations of time or resources.

Attributes of a particular line of scientific inquiry within any discipline can be associated with higher or lower rates of non-replicability. Susceptibility to non-replicability depends on

- the complexity of the system under study;
- the number and relationship of variables within the system under study;
- the ability to control the variables;
- levels of noise within the system (or signal to noise ratios);
- a mismatch of scale of the phenomena and the scale at which it can be measured;
- stability across time and space of the underlying principles;
- fidelity of the available measures to the underlying construct at study (e.g., direct versus indirect measurements); and
- the *a priori* probability (pre-experimental plausibility) of the scientific hypothesis.

Unhelpful sources of non-replicability. In some cases, non-replicability is due to shortcomings in the design, conduct, and communication of a study. Whether arising from lack of knowledge, perverse incentives, sloppiness, or bias, these sources of non-replicability reduce the efficiency of scientific progress; time spent resolving non-replicability issues that are found to be caused by these sources is time not spent expanding scientific understanding.

These sources of non-replicability can be minimized through initiatives and practices aimed at improving design and methodology through training and mentoring, repeating experiments before publication, rigorous peer review, utilizing tools for checking analysis and results, and better transparency in reporting. Efforts to minimize avoidable and unhelpful sources of non-replicability warrant continued attention.

Researchers who knowingly use questionable research practices with the intent to deceive are committing misconduct or fraud. It can be difficult in practice to differentiate between honest mistakes and deliberate misconduct because the underlying action may be the same while the intent is not. Scientific misconduct in the form of misrepresentation and fraud is a continuing concern for all of science, even though it accounts for a very small percentage of published scientific papers.

Improving Reproducibility and Replicability in Research

The committee reviewed current and proposed efforts to improve reproducibility and replicability across science. Efforts to strengthen research practices will improve both. Some efforts are primarily focused on computational reproducibility and others are more focused on replicability, although improving one may also improve the other.

Rigorous research practices were important long before reproducibility and replicability emerged as notable issues in science, but the recent

emphasis on transparency in research has brought new attention to these issues. Broad efforts to improve research practices through education and stronger standards are a response to changes in the environment and practice of science, such as the near ubiquity of advanced computation and the globalization of research capabilities and collaborations.

RECOMMENDATION 6-1: All researchers should include a clear, specific, and complete description of how the reported result was reached. Different areas of study or types of inquiry may require different kinds of information.

Reports should include details appropriate for the type of research, including:

- a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study;
- a clear description of the analysis of data and decisions for exclusion of some data and inclusion of other;
- for results that depend on statistical inference, a description of the analytic decisions and when these decisions were made and whether the study is exploratory or confirmatory;
- a discussion of the expected constraints on generality, such as which methodological features the authors think could be varied without affecting the result and which must remain constant;
- reporting of precision or statistical power; and
- a discussion of the uncertainty of the measurements, results, and inferences.

RECOMMENDATION 6-2: Academic institutions and institutions managing scientific work such as industry and the national laboratories should include training in the proper use of statistical analysis and inference. Researchers who use statistical inference analyses should learn to use them properly.

Improving reproducibility will require efforts by researchers to more completely report their methods, data, and results, and actions by multiple stakeholders across the research enterprise, including educational institutions, funding agencies and organizations, and journals. One area where improvements are needed is in education and training. The use of data and computation is evolving, and the ubiquity of research aided by computation is such that a competent scientist today needs a sophisticated understanding of computation. While researchers want and need to use these tools and methods, their education and training have often not prepared them to do so.

RECOMMENDATION 6-3: Funding agencies and organizations should consider investing in research and development of open-source, usable tools and infrastructure that support reproducibility for a broad range of studies across different domains in a seamless fashion. Concurrently, investments would be helpful in outreach to inform and train researchers on best practices and how to use these tools.

The scholarly record includes many types of objects that underlie a scientific study, including data and code. Ensuring the availability of the complete scholarly record in digital form presents new challenges, including establishing links between related digital objects, making decisions on longevity of storage or access, and enabling the use of stored objects through improved discovery tools (e.g., searches). Many journals and funders do not currently enforce policies to improve the coherence and completeness of objects that are part of the scholarly record.

RECOMMENDATION 6-4: Journals should consider ways to ensure computational reproducibility for publications that make claims based on computations, to the extent ethically and legally possible. Although ensuring such reproducibility prior to publication presents technological and practical challenges for researchers and journals, new tools might make this goal more realistic. Journals should make every reasonable effort to use these tools, make clear and enforce their transparency requirements, and increase the reproducibility of their published articles.

RECOMMENDATION 6-5: In order to facilitate the transparent sharing and availability of digital artifacts, such as data and code, for its studies, the National Science Foundation (NSF) should

- develop a set of criteria for trusted open repositories to be used by the scientific community for objects of the scholarly record;
- seek to harmonize with other funding agencies the repository criteria and data management plans for scholarly objects;
- endorse or consider creating code and data repositories for long-term archiving and preservation of digital artifacts that support claims made in the scholarly record based on NSF-funded research. These archives could be based at the institutional level or be part of, and harmonized with, the NSF-funded Public Access Repository;
- consider extending NSF's current data management plan to include other digital artifacts, such as software; and
- work with communities reliant on nonpublic data or code to develop alternative mechanisms for demonstrating reproducibility.

Through these repository criteria, NSF would enable discoverability and standards for digital scholarly objects and discourage an undue proliferation of repositories, perhaps through endorsing or providing one go-to website that could access NSF-approved repositories.

RECOMMENDATION 6-6: Many stakeholders have a role to play in improving computational reproducibility, including educational institutions, professional societies, researchers, and funders.

- Educational institutions should educate and train students and faculty about computational methods and tools to improve the quality of data and code and to produce reproducible research.
- Professional societies should take responsibility for educating the public and their professional members about the importance and limitations of computational research. Societies have an important role in educating the public about the evolving nature of science and the tools and methods that are used.
- Researchers should collaborate with expert colleagues when their education and training are not adequate to meet the computational requirements of their research.
- In line with its priority for “harnessing the data revolution,” the National Science Foundation (and other funders) should consider funding of activities to promote computational reproducibility.

The costs and resources required to support computational reproducibility for all of science are not known. With respect to previously completed studies, retroactively ensuring computational reproducibility may be prohibitively costly in time and resources. As new computational tools become available to trace and record data, code, and analytic steps, and as the cost of massive digital storage continues to decline, the ideal of computational reproducibility for science may become more affordable, feasible, and routine in the conduct of scientific research.

As with reproducibility, efforts to improve replicability need to be undertaken by individual researchers as well as multiple stakeholders in the research enterprise. Different stakeholders can leverage change in different ways. For example, journals can set publication requirements, and funders can make funding contingent on researchers following certain practices.

RECOMMENDATION 6-7: Journals and scientific societies requesting submissions for conferences should disclose their policies relevant to achieving reproducibility and replicability. The strength of the claims made in a journal article or conference submission should reflect the reproducibility and replicability standards to which an article is held,

with stronger claims reserved for higher expected levels of reproducibility and replicability. Journals and conference organizers are encouraged to:

- set and implement desired standards of reproducibility and replicability and make this one of their priorities, such as deciding which level they wish to achieve for each Transparency and Openness Promotion guideline and working toward that goal;
- adopt policies to reduce the likelihood of non-replicability, such as considering incentives or requirements for research materials transparency, design, and analysis plan transparency, enhanced review of statistical methods, study or analysis plan preregistration, and replication studies; and
- require as a review criterion that all research reports include a thoughtful discussion of the uncertainty in measurements and conclusions.

RECOMMENDATION 6-8: Many considerations enter into decisions about what types of scientific studies to fund, including striking a balance between exploratory and confirmatory research. If private or public funders choose to invest in initiatives on reproducibility and replication, two areas may benefit from additional funding:

- education and training initiatives to ensure that researchers have the knowledge, skills, and tools needed to conduct research in ways that adhere to the highest scientific standards; describe methods clearly, specifically, and completely; and express accurately and appropriately the uncertainty involved in the research; and
- reviews of published work, such as testing the reproducibility of published research, conducting rigorous replication studies, and publishing sound critical commentaries.

RECOMMENDATION 6-9: Funders should require a thoughtful discussion in grant applications of how uncertainties will be evaluated, along with any relevant issues regarding replicability and computational reproducibility. Funders should introduce review of reproducibility and replicability guidelines and activities into their merit-review criteria, as a low-cost way to enhance both.

The tradeoff between resources allocated to exploratory and confirmatory research depends on the field of research, goals of the scientist, mission and goals of the funding agency, and current state of knowledge within a field of study. Exploratory research is more susceptible to non-replication,

while confirmatory research is less likely to uncover exciting new discoveries. Both types of research help move science forward.

RECOMMENDATION 6-10: When funders, researchers, and other stakeholders are considering whether and where to direct resources for replication studies, they should consider the following criteria:

- The scientific results are important for individual decision making or for policy decisions.
- The results have the potential to make a large contribution to basic scientific knowledge.
- The original result is particularly surprising, that is, it is unexpected in light of previous evidence and knowledge.
- There is controversy about the topic.
- There was potential bias in the original investigation, due, for example, to the source of funding.
- There was a weakness or flaw in the design, methods, or analysis of the original study.
- The cost of a replication is offset by the potential value in reaffirming the original results.
- Future expensive and important studies will build on the original scientific results.

CONFIDENCE IN SCIENCE

Replicability and reproducibility are crucial pathways to attaining confidence in scientific knowledge, although not the only ones. Multiple channels of evidence from a variety of studies provide a robust means for gaining confidence in scientific knowledge over time. Research synthesis and meta-analysis, for example, are other widely accepted and practiced methods for assessing the reliability and validity of bodies of research. Studies of ephemeral phenomena, for which direct replications may be impossible, rely on careful characterization of uncertainties and relationships, data from past events, confirmation of models, curation of datasets, and data requirements to justify research decisions and to support scientific results. Despite the inability to replicate or reproduce results of studies of ephemeral phenomena, scientists have made discoveries and continue to expand knowledge of star formation, epidemics, earthquakes, weather, formation of the early universe, and more by following a rigorous process of gathering and analyzing data.

A goal of science is to understand the overall effect from a set of scientific studies, not to strictly determine whether any one study has replicated

any other. Further development in and use of meta-research—that is, the study of research practices—would facilitate learning from scientific studies.

The committee was asked to “consider if the lack of replicability and reproducibility impacts . . . the public’s perception” of science. The committee examined public understanding of science in four relevant areas: factual knowledge, understanding of the scientific process, awareness of scientific consensus, and understanding of uncertainty. Based on evidence from well-designed and long-standing surveys of public perceptions, the public largely trusts scientists. Understanding of the scientific process and methods has remained stable over time, though it is not widespread. NSF’s most recent Science & Engineering Indicators survey shows that 51 percent of Americans understand the logic of experiments and only 23 percent understand the idea of a scientific study.

The committee was not aware of data that would indicate whether there is any link between public perception of science and the lack of replication and reproducibility. The purported existence of a replication “crisis” has been reported in several high-profile articles in mainstream media; however, coverage in public media remains low, and it is unclear whether this issue has registered very deeply with the general population. Nevertheless, scientists and journalists bear responsibility for misrepresentation in the public’s eye when they overstate the implications of scientific research. Finally, individuals and policy makers have a role to play.

RECOMMENDATION 7-1: Scientists should take care to avoid overstating the implications of their research and also exercise caution in their review of press releases, especially when the results bear directly on matters of keen public interest and possible action.

RECOMMENDATION 7-2: Journalists should report on scientific results with as much context and nuance as the medium allows. In covering issues related to replicability and reproducibility, journalists should help their audiences understand the differences between non-reproducibility and non-replicability due to fraudulent conduct of science and instances in which the failure to reproduce or replicate may be due to evolving best practices in methods or inherent uncertainty in science. Particular care in reporting on scientific results is warranted when:

- the scientific system under study is complex and with limited control over alternative explanations or confounding influences;
- a result is particularly surprising or at odds with existing bodies of research;

- the study deals with an emerging area of science that is characterized by significant disagreement or contradictory results within the scientific community; and
- research involves potential conflicts of interest, such as work funded by advocacy groups, affected industry, or others with a stake in the outcomes.

RECOMMENDATION 7-3: Anyone making personal or policy decisions based on scientific evidence should be wary of making a serious decision based on the results, no matter how promising, of a single study. Similarly, no one should take a new, single contrary study as refutation of scientific conclusions supported by multiple lines of previous evidence.

Scientific theories are tested every time someone makes an observation or conducts an experiment, so it is misleading to think of science as an edifice, built on foundations. Rather, scientific knowledge is more like a web. The difference couldn't be more crucial. A tall edifice can collapse—if the foundations upon which it was built turn out to be shaky. But a web can be torn in several parts without causing the collapse of the whole. The damaged threads can be patiently replaced and re-connected with the rest—and the whole web can become stronger, and more intricate.

Nonsense on Stilts: How to Tell Science from Bunk, Massimo Pigliucci